

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effect of sex on specialty training application outcomes: a longitudinal administrative data study of United Kingdom medical graduates
AUTHORS	Woolf, Katherine; Jayaweera, Hirosha; Unwin, Emily; Keshwani, Karim; Valerio, Christopher; Potts, Henry

VERSION 1 - REVIEW

REVIEWER	Dr Tom Gale Peninsula Medical School, Faculty of Medicine and Dentistry, University of Plymouth United Kingdom
REVIEW RETURNED	23-Jul-2018

GENERAL COMMENTS	<p>Well conducted study on large national dataset which has highlighted important findings in relation to sex segregation in certain medical specialties. This research has helped to explain why there are sex differences in specialty training recruitment outcomes due to differential application rates and subsequent success rates in selection processes.</p> <p>Strengths and limitations section: last bullet point should read: large number of tests increased the likelihood of 'type 1 error'</p> <p>Methods, page 5, line 11. It is not clear which years were combined for the analyses. Is this the different student cohorts who entered medical school in 2007 and 2008?</p> <p>Statistical analysis Significant univariate variations were found between medical schools and so medical school attended was included as a random effect in the multivariable models. However, the results tables (3 - 5) do not indicate whether medical school remained significant after adjusting for the other (fixed) factors in the models.</p> <p>The multiple imputation indicators do not include 'medical school attended' and given the between-school differences could this have biased the imputation and hence the final results?</p> <p>A seven-point measure of socio-economic status was created by the authors, by summing six binary variables. Since the authors are treating these variables as measuring an underlying unidimensional construct perhaps they could reference previous studies which have used this same construct or provide some justification regarding the choice of weighting (e.g. calculation of Cronbach's alpha). Also, the SES measure is not mentioned in the</p>
-------------------------	--

	<p>paragraph concerning imputation so how was missing data in the six binary variables handled?</p> <p>The percentages of applied, offered and accepted given in Table 2 would indicate that the denominators used are all doctors, doctors who applied and doctors receiving offers respectively, though this is not explicitly stated. Has the same logic been used in determining the samples used for each of the logistic regression models reported in tables 3, 4 and 5? If so, it would be informative to know the sample sizes for each regression model.</p> <p>Discussion page 11, line 33: anaesthesia should be mentioned as a specialty where men were more likely to apply. page 11, line 36: 'women were more likely to accept an offer versus no offer' does not make sense since they can't accept a 'no offer'.</p> <p>Conclusions It would be good to highlight that the sex differences in general practice which favoured women at each stage of application, offer and acceptance were independent effects. i.e. even though more women applied to GP; within the sample of applicants who applied, more women were offered a post and then more women went on to accept a post. Therefore future efforts which are made to address gender inequality in GP training programmes, should investigate why these differences exist at each stage since this will have a cumulative effect on the gender balance amongst trainees within the specialty.</p>
--	--

REVIEWER	Jason Park MD MEd FRCSC FACS University of Manitoba, Winnipeg, Manitoba, Canada
REVIEW RETURNED	03-Aug-2018

GENERAL COMMENTS	<p>Thank you for the opportunity to review this paper. I enjoyed reading it. My comments are below.</p> <p>This study examines a very large cohort of doctors who applied for a UK specialty position in 2015. I like that the authors controlled for previous academic achievement.</p> <p>1. The title implies that the author are examining whether "professionalism" is associated with application outcomes, but the variable used for this is a Fitness to Practice declaration... is that correct? In most jurisdictions, these document legal cases or unprofessional acts or formally documented serious clinical issues that are a level beyond what I think is useful and applicable to assessing the professional behaviours of most trainees. It takes a complex and difficult to measure construct and reduces it a binary variable. Should readers assume that all those who declare fitness to practice to be totally fine from a professional standpoint? It might reflect unprofessional past behaviours. I don't think it's useful to distinguish grades of professionalism beyond that. I think it's fine to assess this variable. I think it's overstating things to imply that it's a good measure of professionalism and to focus on that as part of your title.</p>
-------------------------	---

	<p>2. What do the authors mean that they included only "first applications"? What happened when trainees applied to more than one program? How did the authors decide? This is not a minor limitation.</p> <p>3. Why wasn't age included as a variable? Could it be a confounder?</p> <p>4. Did the authors consider any statistical adjustments given the number of tests?</p> <p>5. Page 8, line 5: I don't think they are statistically "equivalent". They are statistically "not different". Is that correct?</p> <p>6. My main concern is that the authors present the differences particularly as they relate to women who apply to GP and roll application, offers, acceptances almost together - certainly in how they are presented at parts of the discussion, and again in the conclusion - without really exploring these processes, which are fundamentally different constructs. Applications involve desires, interests, what someone has been encouraged to do (or not discouraged from doing), etc. Offers are different. They involve what programs see as the best candidates or who they want in their programs. These are different constructs, and really need a more thorough and thoughtful discussion in the Discussion section. The processes involved in one are highly unlikely to be the same as in the other.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Dr Tom Gale

Well conducted study on large national dataset which has highlighted important findings in relation to sex segregation in certain medical specialties. This research has helped to explain why there are sex differences in specialty training recruitment outcomes due to differential application rates and subsequent success rates in selection processes.

Strengths and limitations section:

last bullet point should read: large number of tests increased the likelihood of 'type 1 error'

>Thank you for spotting this error. We've corrected it everywhere that it appears.

Methods, page 5, line 11. It is not clear which years were combined for the analyses. Is this the different student cohorts who entered medical school in 2007 and 2008?

>Yes, we combined data on everyone who entered medical school in 2007 and 2008. We have now made this clearer.

Statistical analysis

Significant univariate variations were found between medical schools and so medical school attended was included as a random effect in the multivariable models. However, the results tables (3 - 5) do not indicate whether medical school remained significant after adjusting for the other (fixed) factors in the models.

>We have now included this in the table headings.

The multiple imputation indicators do not include 'medical school attended' and given the between-school differences could this have biased the imputation and hence the final results?

>To examine the impact of including medical school in the imputation model we redid the imputation model for the outcome Applied vs not applied with and without medical school. Please see the table below for the mean values of the imputed variables for each model, which show that including medical school made only a small amount of difference. We did however include medical school as auxiliary variable in the imputation models in the revised version of this MS, as indicated in the revised Statistical Analysis section.

	Original data	25 imputations no med school	25 imputations with med school
UKCAT total	2504.29	2504.37	2503.81
HESA tariff	473.03	440.24	434.75
FPAS SJT	40.77	40.66	40.64
FPAS EPM	38.69	38.70	38.69
FPAS degree	0.942	1.003	0.998
FPAS pub	0.264	0.269	0.269

A seven-point measure of socio-economic status was created by the authors, by summing six binary variables. Since the authors are treating these variables as measuring an underlying unidimensional construct perhaps they could reference previous studies which have used this same construct or provide some justification regarding the choice of weighting (e.g. calculation of Cronbach's alpha). Also, the SES measure is not mentioned in the paragraph concerning imputation so how was missing data in the six binary variables handled?

>We thank Dr Gale for his useful comments and questions on this point. SES was not the main focus of this study. We had included it only as a potential confounder and thus had felt that including one combined variable was therefore more appropriate than including all six SES variables.

When we calculated Cronbach's alpha for the combined variable as suggested, it was 0.53. Furthermore correlations between most of the items were low, with the exception of Free School Meals and Income Support, which were correlated at $r=.56$, and socio-economic class (parental occupation) and parental degree, which were correlated at $r=.34$.

We have therefore redone the regression analyses using the binary socioeconomic class variable (based on parental occupation), school type, POLAR3, and free school meals as measures of socio-economic status. We chose those variables because they measure slightly different things: POLAR3 is the only area level variable, socio-economic class (SEC) is a measure of parental occupation, school type is measure of the student's educational background (and students from poor backgrounds can receive scholarships to attend private schools), and free school meals is a marker of parental income. Furthermore they had less missingness (SEC=8.7%); school type (8.7%); POLAR3 (9.7%) compared to parental education (13.8%) and income support (19.9%). Free school meals was included despite having 16.3% missingness to ensure there was a measure closely related to income.

We had not previously imputed missing values for the SEC variable, meaning cases with missing values were deleted listwise from the regression models. We have now however imputed the missing values for the four socio-economic variables, and recalculated the regression models. This is described in the revised Statistical Analysis section.

It has not changed the overall results much.

The percentages of applied, offered and accepted given in Table 2 would indicate that the denominators used are all doctors, doctors who applied and doctors receiving offers respectively, though this is not explicitly stated. Has the same logic been used in determining the samples used for each of the logistic regression models reported in tables 3, 4 and 5? If so, it would be informative to know the sample sizes for each regression model.

>Yes, this is correct, and we have added this in to the titles of Tables 3, 4 and 5.

Discussion

page 11, line 33: anaesthesia should be mentioned as a specialty where men were more likely to apply.

>Thank you, we have added this in now.

page 11, line 36: 'women were more likely to accept an offer versus no offer' does not make sense since they can't accept a 'no offer'.

>We have reworded this to "women were more likely to accept an offer if they received one".

Conclusions

It would be good to highlight that the sex differences in general practice which favoured women at each stage of application, offer and acceptance were independent effects. i.e. even though more women applied to GP; within the sample of applicants who applied, more women were offered a post and then more women went on to accept a post. Therefore future efforts which are made to address gender inequality in GP training programmes, should investigate why these differences exist at each stage since this will have a cumulative effect on the gender balance amongst trainees within the specialty.

>We have expanded the discussion and are now more explicit about differences in applications and acceptances being the result of choices made by doctors, whereas differences in offers being the result of the selection processes. We explain that there may have been sex differences on the assessments used in GP selection which could explain the sex differences in offers. We have also expanded the conclusion to make this clearer, and have also added a few words to the Conclusion of the abstract to make it clearer that sex differences in acceptances are among those receiving an offer.

Reviewer: 2

Reviewer Name: Jason Park MD MEd FRCSC FACS

Institution and Country: University of Manitoba, Winnipeg, Manitoba, Canada

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Thank you for the opportunity to review this paper. I enjoyed reading it. My comments are below.

This study examines a very large cohort of doctors who applied for a UK specialty position in 2015. I like that the authors controlled for previous academic achievement.

1. The title implies that the author are examining whether "professionalism" is associated with application outcomes, but the variable used for this is a Fitness to Practice declaration... is that correct? In most jurisdictions, these document legal cases or unprofessional acts or formally documented serious clinical issues that are a level beyond what I think is useful and applicable to assessing the professional behaviours of most trainees. It takes a complex and difficult to measure construct and reduces it a binary variable. Should readers assume that all those who declare fitness to practice to be totally fine from a professional standpoint? It might reflect unprofessional past behaviours. I don't think it's useful to distinguish grades of professionalism beyond that. I think it's fine to assess this variable. I think it's overstating things to imply that it's a good measure of professionalism and to focus on that as part of your title.

>We agree that whether or not a doctor declared having had a fitness to practise issue at medical school is unlikely to capture all aspects of a trainee doctor's professionalism. However, previous research has shown that having a professionalism issue at medical school is a predictor of

subsequent disciplinary action (Papadakis, Hodgson, Teherani, Kohatsu, 2004) and as such we suggest that it is not *un*meaningful. It was also the only measure of professionalism available to us. We do agree however that its inclusion in the title may be an overstatement and have therefore removed it.

We should also say here for completeness that when we re-did the analyses (see above), there was a small association between having an FtP declaration and being less likely to apply to specialty training (vs applying).

2. What do the authors mean that they included only "first applications"? What happened when trainees applied to more than one program? How did the authors decide? This is not a minor limitation.

>We analysed applications to specialties separately. As such, a doctor who applied to, say General Practice and Anaesthetics, would have been counted in the analysis of applications, offers and (if applicable) offers to GP and in the same analyses for Anaesthetics. We have made this more explicit in the methods. We agree this is important, and have emphasised it to the reader by putting it in the Strengths and Limitations section.

3. Why wasn't age included as a variable? Could it be a confounder?

>We have now included year of birth as a variable in the analyses.

4. Did the authors consider any statistical adjustments given the number of tests?

>We have used asterisks to highlight where results were significant at $p < .001$, and have emphasised (e.g. including by putting in the Strengths and Limitations section) that the reader should consider the number of tests in interpreting the findings. We also explicitly mention this in the summary of findings in the discussion: "... although the large number of tests performed increases the possibility that this was due to type I error."

5. Page 8, line 5: I don't think they are statistically "equivalent". They are statistically "not different". Is that correct?

>Thank you – we have changed this in the text.

6. My main concern is that the authors present the differences particularly as they relate to women who apply to GP and roll application, offers, acceptances almost together - certainly in how they are presented at parts of the discussion, and again in the conclusion - without really exploring these processes, which are fundamentally different constructs. Applications involve desires, interests, what someone has been encouraged to do (or not discouraged from doing), etc. Offers are different. They involve what programs see as the best candidates or who they want in their programs. These are different constructs, and really need a more thorough and thoughtful discussion in the Discussion section. The processes involved in one are highly unlikely to be the same as in the other.

>As described in the response to Reviewer 1, we have now expanded the discussion and the conclusion to include these important points.

VERSION 2 – REVIEW

REVIEWER	Tom Gale University of Plymouth United Kingdom
REVIEW RETURNED	18-Jan-2019

GENERAL COMMENTS	The authors have addressed the concerns I had regarding the clarification of analyses undertaken and have expanded the
-------------------------	--

	<p>discussion significantly to highlight implications for policy and practice, with respect to workforce planning and recruitment. This study will receive significant international interest from a wide array of people involved with postgraduate training programmes and workforce planning.</p>
--	--